doi.org.10.54473/IJTRET.2023.7302
International Journal of Trendy Research in Engineering and Technology
Volume 7 Issue 3 June 2023
ISSN NO 2582-0958

IDENTIFYING FREQUENT POINT OF INTEREST USING GEOTAGGED PHOTOS

Keerthipriya , A.Gulsan Banu
Department of Computer Science Engineering,
Sree Sakthi Engineering College,
Karamadai, Mettupalayam, India
selvakeerthi17@gmail.com, gulshanasif97@gmail.com

Received 09 May 2023 Received in revised form 12 May 2023 Accepted 13 May 2023
Available online 19 May 2023

ABSTRACT

Tourist Recommending has been a growing interest and it is identified based on the Relevant Point of Interest (POIs) to form a Personalized Trip using Location-Based Social Networks database. Most frequently visited places are identifiedas person's point of interest. The aim of this paper is to identify and analyze the main tourist attraction places in a city. In this work Association rule mining Algorithm is used to analyze the frequently visited locations along with the visitor's information. For the analysis R programming language is used. The dataset Flickr is used to identify the places where most frequently visited based on geospatial information. The Current work read the Flickr dataset. The dataset is divided into different segments and individual segments are analyzed. Final result is generated by combining the results of the individual segments.

Keywords: location based social network, trip recommendation, and spatiotemporal data mining.

I. INTRODUCTION

This research paper deals with the interests of tourists and gives an overview of the places that are frequently visited by tourists. It is difficult for a tourist to know about the places he visits in a city. The behaviour and semantic annotation can be identified using point of interest (POI). The advancement in web-based and mobile technologies gives users the ability to access social media, which in turn provides more data for analysis. Social media is user-created online content that allows users to interact and communicate with their friends, colleagues, and families. On the web, photos can be shared through popular photo-sharing websites such as Flickr. Flickr is the most popular photo sharing website. The association rule is the most popular and well-known approach for discovering interesting relationships between variables in large databases. The arules package, available in r, provides methods for creating and manipulating input datasets. And also analysis of the resulting data sets and rules. R programming is a statistical programming language useful for analysing statistical problems and all other related data mining algorithms. This algorithm can be used to find frequent item sets in a given city.

II. LITERATURE SURVEY

Xiaoting Wang et al[1] identified the travel recommendation with the relevant points of interest (POIs) and also selected the best time of day to visit the POIs. They propose the personalized crowd-aware trip recommendation (PersCT) algorithm to recommend personalized trips that avoid the most crowded times of POIs. This algorithm evaluates pedestrian traffic information from a real pedestrian sensor dataset and user travel history, where PersCT uses real datasets from geo-tagged Flickr photos. Finally, the PersCT method displays the entire travel history information .

Ickjai Lee et al [2] identifies points of interest and their associations using data mining techniques such asclustering and association rules to identify points of attraction and their association patterns. For the analysis, data is collected from flickr in the Queensland, Australia, area. The most popular travel destinations are located there. They propose a PoI association mining framework for geotagged photos using a combination of two popular data mining techniques, clustering and association rule mining. They focus on Queensland, one of the most popular travel destinations in Australia.

Ickjai Lee et al [3] in their paper 'Points-of-Interest Mining from People's Photo- Taking Behavior', discusses with point of interest mining. People's behavior can be identified based on their interesting photo patterns. Ickjai Lee analyzes geotagged photos from flickr for Queensland and Australia. Based on people's photo patterns, the most interesting places are used for decision making by local businesses, policy makers.

Thanh-Hieu Bui et al [4] discusses point of interest (POI) mining from a collection of geotagged photos combined with appropriate semantic annotation using additional POI information from external POI databases with high coverage. The POI mining refers to the processes of pattern recognition, i.e. clustering, extraction and semantic annotation. The author proposes a novel POI mining framework using two-stage clustering, random walk and constrained clustering. Experimental results on two datasets of geotagged flickr photos of two cities in California, USA. Finally, they compare the existing approaches and propose a novel clustering framework based on proper semantic annotation using additional POI information from an external POI database with high coverage.

Cai et al[5] discussed the identification of photographers' most interesting regions of interest (RoI) and most frequent trajectory patterns using the trajectory pattern mining algorithm with a dataset from Flickr. The software used in the work is the programming language JAVA. The trajectory pattern mining algorithm (TPM) requires three user-supplied input parameters, for example, to find out the region of interest and frequent patterns with the Flickr dataset, which contains different spatial regions: Australia Queensland. Using the TPM algorithm, the photos from the above regions were analyzed and are able to identify major cities, tourist spots, and expected tourist routes without external user input.

Zheng et al [6] , inform that a region of attraction is a place of interest visited by tourists one after another. With the help of GPS tagged photos, it is easy to identify most points of attraction. The typical itinerary and region of attraction (RoA) using GPS tagged photos, GPS -tagged photos downloaded from the Internet, contain the local travel database, build a statistically reliable database of itineraries, and create a list of regions of attraction (RoA). Then, the database of travel routes will be analyzed using the entropy- based mobility measure and Z-test. Using DBSCAN density-based clustering algorithm can generate the regions of attractions (RoA) and able to identify the four major cities, it includes San Francisco, New York City, Paris and London.

III. IMPLEMENTATION DETAILS

A point of interest is identified as a particular city that the user frequently visits. In this work, it is created using an association rule mining algorithm from the R package. The R package arules presented in this work allows the creation and manipulation of input datasets and the analysis of the resulting itemsets and rules. The dataset used for the analyzes is a flickr dataset that includes a set of users and their visits to different locations in the city, with a total of 3975 tours and 17,087 visits. The visits of the user POI are determined based on the geotagged YFCC100M Flickr photos. This dataset contains userid, photoid, date taken, PoiId, Poi topic, SeqId, latitude and longitude. Association rule mining can be used to identify the frequent location. Association rule is the most popular and researched approach for detecting interesting relationships between variables in large databases.

R is a powerful language and environment for statistical computing and graphics. R is open source and is widely used by statisticians, biostatisticians, and geneticists. There is a wealth of libraries that can be built directly into our code. Mining association rules is a popular mining approach for discovering positive associations. Given a set I = {I1, I2, ..., Ik} of entries (k-itemsets) in a transaction database D. Each transaction T ∈ D is a subset of I. This method is often referred to as frequent pattern mining because it discovers frequent patterns. A k-itemset is frequent if its frequency is greater than or equal to a user-specified threshold. An association rule is an expression in the form X = > Y ($X \cap Y = \varphi$), where X is antecedent and Y is consequent.

Two estimates are required in association rule discovery: support and confidence. The support of an itemset X ∈ D is the number of transactions in D that contain X. That is, it is the probability that X and Y occur in the dataset. The confidence of an association rule X = > Y is the conditional probability that Y is in a transaction if X is in that transaction. Support = probability (X ∪ Y), confidence = probability (X ∪ Y)/probability (X). For the current analysis, the support value is 0.01 and the confidence value is 0.5 to find an association between elements. The arules package is used for mining association rules and frequent itemsets using the R programming language. The arules package for R provides the infrastructure for representing, manipulating, and analyzing transactional data and patterns (frequent itemsets and association rules). Mining associations with Apriori is used to mine frequent itemsets, association rules, or association hyperedges. The Apriori algorithm uses a stepwise search for frequent itemsets.

International Journal of Trendy Research in Engineering and Technology
Volume 7 Issue 3 June 2023
ISSN NO 2582-0958

Table 1: User Visit Data

photoID	userID	date/time taken	poiID	poiTheme	poiFreq	seqID
12344732513	100895643	1/27/2014	120	Leisure/R	357	1
12344760173	100895643	1/27/2014	188	Place of W	225	1
12321530175	100895643	2/5/2014	190	Transport	351	2
12321735773	100895643	2/5/2014	120	Leisure/R	357	2
12322029584	100895643	2/5/2014	120	Leisure/R	357	2
12322044814	100895643	2/5/2014	120	Leisure/R	357	2
11768725924	101884347	12/13/2013	97	Office	148	3
11769094426	101884347	12/13/2013	97	Office	148	3
11768756124	101884347	12/13/2013	189	Place Of A	765	3
11768597463	101884347	12/13/2013	189	Place Of A	765	3
11768762064	101884347	12/13/2013	189	Place Of A	765	3
11768766564	101884347	12/13/2013	189	Place Of A	765	3
11768359575	101884347	12/13/2013	189	Place Of A	765	3
11768616573	101884347	12/13/2013	189	Place Of A	765	3
11768619193	101884347	12/13/2013	189	Place Of A	765	3
11768378535	101884347	12/13/2013	189	Place Of A	765	3
11768382115	101884347	12/13/2013	189	Place Of A	765	3
11768636943	101884347	12/13/2013	189	Place Of A	765	3
11768643743	101884347	12/13/2013	189	Place Of A	765	3
909465863	10195518(	6/3/2007	57	Place Of A	58	4
1145062626	10195518(	8/15/2007	6	Place of W	67	5
1144222893	10195518(	8/15/2007	6	Place of W	67	5
1144228037	10195518(	8/15/2007	241	Place Of A	23	5

The table 1 depicts the user visit data that contains the relevant data such that photoid, userid, date/time taken, poiid, poitheme, poifrequency, sequenceid.

Table 2: Geospatial Data

poiID	theme	subTheme	poiName	lat	long
1	Transport	Railway St	Flemingtc	-37.7882	144.9393
2	Mixed Use	Retail/Off	Council Hc	-37.8143	144.9666
3	Place Of A	Library	The Melbc	-37.8149	144.9673
4	Leisure/R	Informal C	Carlton Ga	-37.8061	144.9713
5	Place of W	Church	St Francis	-37.8119	144.9624
6	Place of W	Church	Wesley Ch	-37.8102	144.9682
7	Place of W	Church	St Augusti	-37.817	144.9549
8	Place of W	Church	St James C	-37.8101	144.9525
9	Place of W	Church	St Mary's A	-37.8032	144.9538
10	Place of W	Church	Romanian	-37.8052	144.967
11	Place of W	Church	Welsh Pre	-37.8104	144.9599
12	Place of W	Church	Church of	-37.8105	144.9639
13	Place of W	Church	Scots Chu	-37.8146	144.9686
14	Place of W	Church	St Michae	-37.8144	144.9692
15	Place of W	Church	Greek Ort	-37.8088	144.9783
16	Place of W	Church	St Peter's	-37.8097	144.9753
17	Place of W	Church	Lutheran T	-37.811	144.9757
18	Place of W	Church	Holy Trinit	-37.8141	144.9832
19	Place of W	Church	St Johns Li	-37.8209	144.9671
20	Place of W	Church	North Mel	-37.8036	144.9477
21	Place of W	Church	Melbourn	-37.8114	144.9847
22	Place of W	Church	All Nation	-37.7959	144.969
23	Place of W	Church	Our Lady c	-37.8026	144.9693
24	Place of W	Church	St Michae	-37.7941	144.9454

The table 2 depicts the geospatial data that contains the relevant information that is photoid, theme, subtheme, poiname, latitude and longitude data.

IV. ALGORITHM

In this work, the Apriori algorithm is used to identify frequent points of interest (POI). Apriori is an algorithm for mining frequent items and learning association rules in transaction databases. Thealgorithm identifies the frequent individual points in the database and expands them to larger and larger point sets as long as these point sets occur sufficiently often in the database. The frequent item sets identified by Apriori can be used to identify association rules that show general trends in the database: This has applications in areas such as shopping cart analysis. 17.088 items are included in the dataset that we split into individual 0f 300
Using the Apriori algorithm, the frequent item set of the transaction is displayed so that frequent points of interest (POI) can be easily identified.
The dataset from flickr was used for the analyzes. The geotagged photos contain spatiotemporal information such as latitude and longitude.

The input for the analyzes is the data from userid and poitheme. These two datasets are used for identifying common elements. Then the final output is displayed with the graph as userid and poi theme value.

Steps involved in the algorithm

Step 1: Reading the data to run an algorithm. To read a file in R, the files are first read into a data frame, which is denoted by data. Then header= TRUE indicates that this data contains a header and sep="," indicates that the data is separated by commas.

Step 2: Split the files into separate files for the execution of data.

Step 3: To view the file in R, the view function is called with the file name and the files are displayed.

Step 4: After splitting, the algorithm is used for analysis. For the Apriori algorithm, the file name with the support value and the confidence value is used for the analyzes.

Step 5: Then the final result is saved in the separate files.
Output for algorithm The five charts display the execution of algorithm with the splitting of 300 datas.

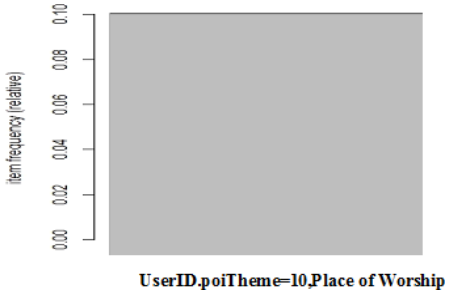

Figure 1: Place of Worship

In this Figure.1 represent the higher value of data with the userid and poitheme.The value of useridis 10 and the poi theme is Place of Worship and the frequency value is 10.

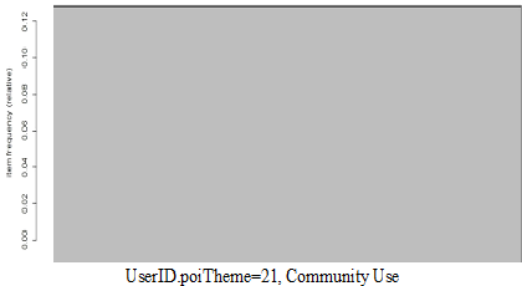

Figure 2: Community Use

In this Figure.2 represent the higher value of data with the userid and poitheme.The value of userid is 21 and the poi theme is Community use and the frequency value is 13.

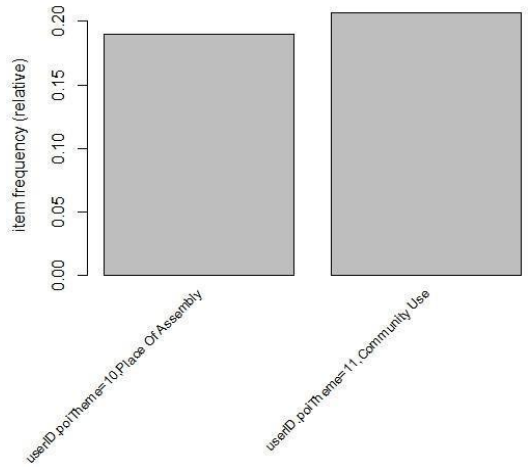

Figure 3: Community Use

In this Figure 3 represent the higher value of data with the userid and poitheme.The value of useridis 10 and the poi theme is Place of Assembly, and then another value is higher userid is 11 and the poi theme is Community use. Then the frequency value of both is 18 and 20.

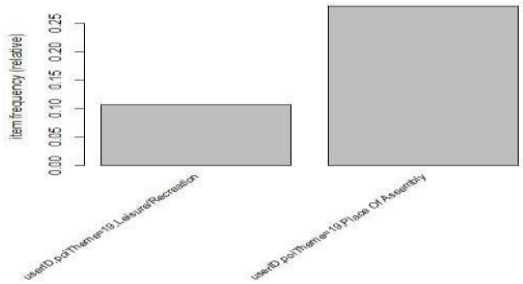

Figure 4: Place of Assembly

In Figure.4 represent the higher value of data with the userid and poitheme.The value of useridis 19 and the poi theme is Leisure Recreation, and then another value is higher userid is 19 and the poi theme is place of Assembly. Then the frequency value of both is 10 and 25.

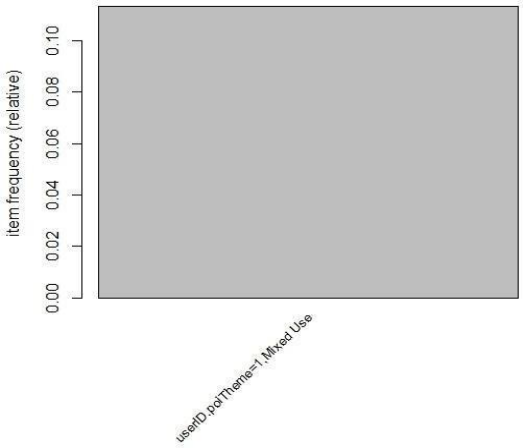

Figure 5: Mixed Use

In this figure .5 represent the higher value of data with the userid and poitheme.The value of userid is 1 and the poi theme is Missed Use. Then the frequency value 14.

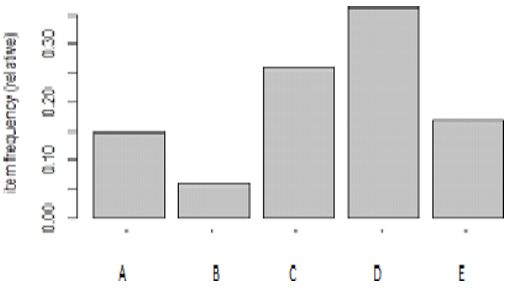

Figure 6: Place of Assembly

The value of a,b,c,d,e are

- A= userid.poitheme=1,mixed use
- B= userid.poitheme=10, Place of Assembly.
- C=userid.poitheme=11, community use.
- D= userid.poitheme=19, Place of Assembly.
- E= userid.poitheme=21,community use .

Finally when combining the five individual chart data and then execute them in the algorithm it display the below chart the most frequent Point of Interest is Place of Assembly. Result shows the place of Assembly is most frequency tourist people visited again and again.

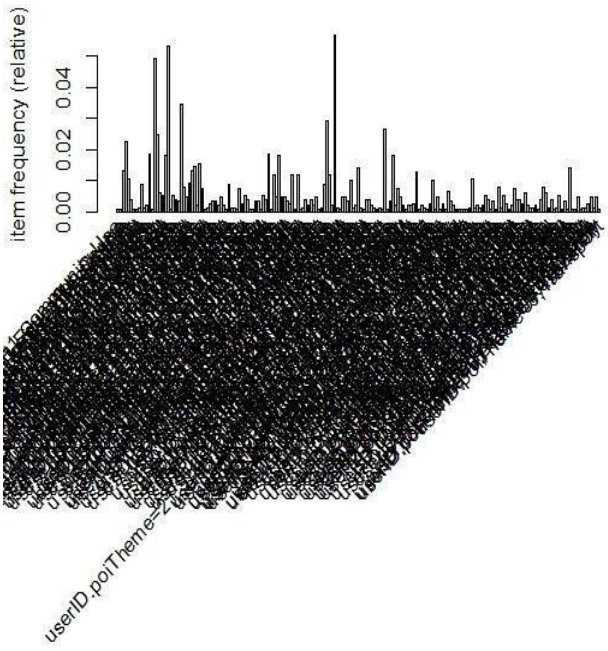

Figure .7: Overall Data Figure

Overall when executing the 17,088 dataset the result shows does not show the clear information of the frequency data. So we split the data and execute them in algorithm it shows the clear information of frequency point of interest.

Our Algorithm gives advantages while we are considering following points:

- Distributing Computing Support.
- Output is same for comprising splitted information and whole they process.
- Viewability and Clarity will be more when we are splitting the input and do the analyzes.
- Comparing to the other figures it will be clarity.

V. CONCLUSION

The geotagged photos are useful for gathering the spatial data information. The user take photos where they frequently go and visit in a particular place and interesting location and post on the flickr are social media of popular photo sharing sites. The photos contain some information of userid and destination location of photo taken. That dataset is used for finding frequent tourist destination of a particular city. With the dataset we use proper algorithm as association rule mining and analyzes the most frequent place of the city in a particular city.

REFERENCES

[1]. Xiaoting Wang, Christopher Leckie, Jeffery Chan, Kwan Hui Lim and Tharshan Vaithianathan."Improving Personalized Trip Recommendation to Avoid Crowds Using Pedestrian Sensor Data". Proceedings of the 25th ACM International Conference on Information and Knowledge Management (CIKM'16),(2016) 25-34. 2016.

[2]. Ickjai Lee, Guochen Cai, Kyungmi Lee" Mining Points-of-Interest Association Rules from Geo-tagged Photos" Proceedings of 46th Hawaii International Conference on System Sciences.(2013)

[3]. Ickjai Lee, Guochen Cai, Kyungmi Lee "Points-of-Interest Mining from People's Photo- Taking Behavior", Proceedings of 46th Hawaii International Conference on System Sciences.(2013)

[4]. Thanh-Hieu Bui & Seong-Bae Park "Point of interest mining with proper semantic annotation", Multimedia Tools and Applications 76 (2017)23435–23457https://doi.org/10.1007/s11042-016-4114-7

[5]. Guochen Cai, Chihiro Hio, Luke Bermingham, Kyungmi Lee, Ickjai Lee, 'Mining Frequent Trajectory Patterns and Regions-of-Interest from Flickr Photos', Proceedings of 47th Hawaii International Conference on System Science ,Australia (2014).

[6]. Zheng, YT., Li, Y., Zha, ZJ., Chua, TS. (2011). Mining Travel Patterns from GPS-Tagged Photos. In: Lee, KT., Tsai, WH., Liao, HY.M., Chen, T., Hsieh, JW., Tseng, CC. (eds) Advances in Multimedia Modeling. MMM 2011. Lecture Notes in Computer Science, vol 6523. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-17832-0_25